# Improving Convergence of Evolutionary Multi-Objective Optimization with Local search - A Concurrent-Hybrid Algorithm.

Karthik Sindhya

Department of Mathematical Information Technology,
Industrial Optimization Group, P.O. Box 35 (Agora),
FI-40014 University of Jyväskylä, Finland

March 26, 2009

# Myself

- Born: Bangalore, India.

# Myself

- Born: Bangalore, India.
- On map

# Myself

- Born: Bangalore, India.
- On map


- Education: Bachelor and Master's in Engineering (Chemical Engineering).
- Doctoral Student in Mechanical Engineering at Kanpur Genetic Algorithms Laboratory, IIT Kanpur.

# Myself

- Visiting Research Student at Helsinki School of Economics.
- Doctoral student at University of Jyväskylä.

# Myself

- Visiting Research Student at Helsinki School of Economics.
- Doctoral student at University of Jyväskylä.
- Research Interests: Evolutionary Algorithms, Evolutionary Multi-objective Optimization, Artificial Neural Networks and Multiple Criteria Decision Making.

# Myself

- Visiting Research Student at Helsinki School of Economics.
- Doctoral student at University of Jyväskylä.
- Research Interests: Evolutionary Algorithms, Evolutionary Multi-objective Optimization, Artificial Neural Networks and Multiple Criteria Decision Making.
- Thesis Advisors:
  - Prof. Kaisa Miettinen, Department of Mathematical Information Technology, University of Jyväskylä, Finland.
  - Prof. Kalyanmoy Deb, Department of Business Technology, Helsinki School of Economics, Finland.
    - Department of Mechanical Engineering, Indian Institute of Technology Kanpur, India.

# Introduction

- Evolutionary algorithm have been a widely used approach to solve multi-objective optimization problems for a decade.

# Introduction

- Evolutionary algorithm have been a widely used approach to solve multi-objective optimization problems for a decade.
- Evolutionary multi-objective optimization (EMO) deals with a population of points and yields a set of solutions which are non-dominated and near Pareto-optimal.
    - Idea is to generate an approximate non-dominated set which represents the Pareto-optimal front.

# Introduction



- Evolutionary algorithm have been a widely used approach to solve multi-objective optimization problems for a decade.
- Evolutionary multi-objective optimization (EMO) deals with a population of points and yields a set of solutions which are non-dominated and near Pareto-optimal.
  - Idea is to generate an approximate non-dominated set which represents the Pareto-optimal front.
- In EMO, there are clearly two important goals:
  - Convergence to the Pareto-optimal front.
  - Diverse set of solutions in the non-dominated front.

# Introduction





- Evolutionary algorithm have been a widely used approach to solve multi-objective optimization problems for a decade.
- Evolutionary multi-objective optimization (EMO) deals with a population of points and yields a set of solutions which are non-dominated and near Pareto-optimal.
  - Idea is to generate an approximate non-dominated set which represents the Pareto-optimal front.
- In EMO, there are clearly two important goals:
  - Convergence to the Pareto-optimal front.
  - Diverse set of solutions in the non-dominated front.
- Main advantages of EMO algorithms:-
  - Obtaining a set of non-dominated solutions in a single run.
  - Ease in handling multiple local Pareto-optimal fronts.
  - Flexibilities in handling of discrete, nonlinear, multi-modal and large-scale problems.

# Introduction



- EMO approaches are often criticized for their lack of theoretical convergence proof to the Pareto-optimal front.

# Introduction



- EMO approaches are often criticized for their lack of theoretical convergence proof to the Pareto-optimal front.
- Multiple criteria decision-making (MCDM) techniques are also commonly used to deal with multi-objective optimization problems.
  - Have theoretical convergence proofs.
  - Multi-objective problem $\rightarrow$ Single-objective problem and solved by any mathematical programming technique.
  - Typically a single Pareto-optimal solution.

# Introduction

- EMO approaches are often criticized for their lack of theoretical convergence proof to the Pareto-optimal front.
- Multiple criteria decision-making (MCDM) techniques are also commonly used to deal with multi-objective optimization problems.
  - Have theoretical convergence proofs.
  - Multi-objective problem $\rightarrow$ Single-objective problem and solved by any mathematical programming technique.
  - Typically a single Pareto-optimal solution.
- EMO criticism can be bridged by incorporating MCDM approaches into EMO.

# Introduction



- EMO approaches are often criticized for their lack of theoretical convergence proof to the Pareto-optimal front.
- Multiple criteria decision-making (MCDM) techniques are also commonly used to deal with multi-objective optimization problems.
  - Have theoretical convergence proofs.
  - Multi-objective problem $\rightarrow$ Single-objective problem and solved by any mathematical programming technique.
  - Typically a single Pareto-optimal solution.
- EMO criticism can be bridged by incorporating MCDM approaches into EMO.
- Integration of MCDM in EMO is not straightforward.

# Introduction

- EMO approaches are often criticized for their lack of theoretical convergence proof to the Pareto-optimal front.
- Multiple criteria decision-making (MCDM) techniques are also commonly used to deal with multi-objective optimization problems.
  - Have theoretical convergence proofs.
  - Multi-objective problem $\rightarrow$ Single-objective problem and solved by any mathematical programming technique.
  - Typically a single Pareto-optimal solution.
- EMO criticism can be bridged by incorporating MCDM approaches into EMO.
- Integration of MCDM in EMO is not straightforward.
- One way: EMO as a global optimizer and MCDM approach as a local optimizer.

# Serial Approach

- Hybrid Algorithms have been broadly classified into **serial** and **concurrent** approaches.


Figure: Serial approach.

- E.g. MSGA-(LS1, LS2, LS3), Goel and Deb etc.,

# Serial Approach

- Hybrid Algorithms have been broadly classified into **serial** and **concurrent** approaches.


Figure: Serial approach.

- E.g. MSGA-(LS1, LS2, LS3), Goel and Deb etc.,
- Adv: Convergence to Pareto-optimal front.

# Serial Approach

- Hybrid Algorithms have been broadly classified into **serial** and **concurrent** approaches.


Figure: Serial approach.

- E.g. MSGA-(LS1, LS2, LS3), Goel and Deb etc.,
- Adv: Convergence to Pareto-optimal front.
- Shortcoming: Switchover from global to local search.

# Concurrent Approach

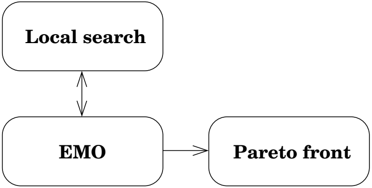


Figure: Concurrent approach.

- E.g. MOGA by Ishibuchi, MOGLS by Jaszkiewicz etc.,

# Concurrent Approach



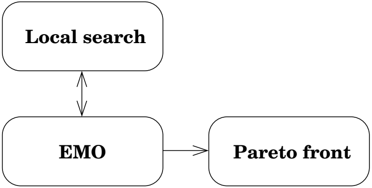


Figure: Concurrent approach.

- E.g. MOGA by Ishibuchi, MOGLS by Jaszkiewicz etc.,
- Advantages:
  - Convergence to Pareto-optimal front.
  - Faster convergence.
  - No switchover problem.

# Concurrent Approach

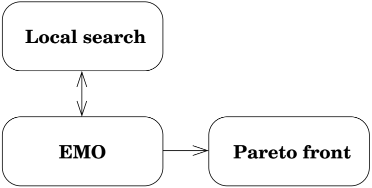


Figure: Concurrent approach.

- E.g. MOGA by Ishibuchi, MOGLS by Jaszkiewicz etc.,
- Advantages:
  - Convergence to Pareto-optimal front.
  - Faster convergence.
  - No switchover problem.
- Shortcoming: Which and frequency of the EMO individuals to be local searched?

# Summary of Literature Survey



- Weighted sum of objectives is the most common scalarizing procedure.

# Summary of Literature Survey



- Weighted sum of objectives is the most common scalarizing procedure.
  - All points on the Pareto-optimal front is impossible unless the Pareto-optimal front is **convex**.

# Summary of Literature Survey

- Weighted sum of objectives is the most common scalarizing procedure.
  - All points on the Pareto-optimal front is impossible unless the Pareto-optimal front is **convex**.
- No clear winner.
  - Every algorithm is applied on a different set of test functions and performance criteria.

# Summary of Literature Survey



- Weighted sum of objectives is the most common scalarizing procedure.
    - All points on the Pareto-optimal front is impossible unless the Pareto-optimal front is **convex**.
- No clear winner.
    - Every algorithm is applied on a different set of test functions and performance criteria.
- We chose **Concurrent approach** and better scalarizing function called **achievement scalarizing function (ASF)**.

# Achievement Scalarizing Function

- We consider a multi-objective optimization problem of the form:

$$\begin{array}{rl} \text{minimize} & \{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots.., f_k(\mathbf{x})\} \\ \text{subject to} & \mathbf{x} \in \mathcal{S}, \end{array} \tag{1}$$

# Achievement Scalarizing Function



- We consider a multi-objective optimization problem of the form:

$$\begin{aligned} \text{minimize} \quad & \{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots.., f_k(\mathbf{x})\} \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{S}, \end{aligned} \tag{1}$$

- An example of an augmented achievement scalarizing function is given by:

$$\begin{aligned} \text{minimize} \quad & \max_{i=1}^{k} \frac{f_i(\mathbf{x}) - \bar{\mathbf{z}}_i}{\mathbf{z}_i^{max} - \mathbf{z}_i^{min}} + \rho \sum_{i=1}^{k} \frac{f_i(\mathbf{x}) - \bar{\mathbf{z}}_i}{\mathbf{z}_i^{max} - \mathbf{z}_i^{min}}, \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{S}, \end{aligned} \tag{2}$$

# Achievement Scalarizing Function



- We consider a multi-objective optimization problem of the form:

$$\begin{aligned} \text{minimize} \quad & \{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots\ldots, f_k(\mathbf{x})\} \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{S}, \end{aligned} \tag{1}$$

- An example of an augmented achievement scalarizing function is given by:

$$\begin{aligned} \text{minimize} \quad & \max_{i=1}^{k} \frac{f_i(\mathbf{x}) - \bar{\mathbf{z}}_i}{\mathbf{z}_i^{max} - \mathbf{z}_i^{min}} + \rho \sum_{i=1}^{k} \frac{f_i(\mathbf{x}) - \bar{\mathbf{z}}_i}{\mathbf{z}_i^{max} - \mathbf{z}_i^{min}}, \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{S}, \end{aligned} \tag{2}$$

  - $\frac{1}{\mathbf{z}_i^{max} - \mathbf{z}_i^{min}}$ is a weight factor assigned to each objective function $f_i$.
  - The weighing factors are used to normalize the values of each objective function $f_i$.
  - $\bar{\mathbf{z}} \in R^k$ is a reference point.
  - $\rho > 0$, binds the trade-offs called an augmentation coefficient.

# Achievement Scalarizing Function



- Advantages:
  - The optimal solution of an ASF is always Pareto-optimal.
  - Any Pareto-optimal solution can be obtained by changing the reference point.
  - The optimal value of an ASF is zero, when the reference point is Pareto-optimal.

# A concurrent-Hybrid Algorithm



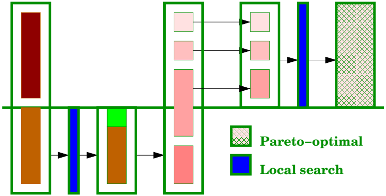


Figure: Concurrent-hybrid algorithm.

# Probability of Local Search-$P_t^{local}$

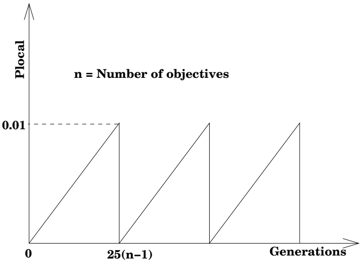


Figure: Probability of local search.

- To maintain exploration-exploitation balance.

# Termination criteria



- Till date EMO algorithms are usually terminated in any of the following ways:
  - A pre-specified number of generations.
  - No new solutions have entered the non-dominated set after a prefixed number of generations.

# Termination criteria



- Till date EMO algorithms are usually terminated in any of the following ways:
  - A pre-specified number of generations.
  - No new solutions have entered the non-dominated set after a prefixed number of generations.
- We utilize the slack variable $\alpha$ for a new convergence criterion.

# Termination criteria



- Till date EMO algorithms are usually terminated in any of the following ways:
  - A pre-specified number of generations.
  - No new solutions have entered the non-dominated set after a prefixed number of generations.
- We utilize the slack variable $\alpha$ for a new convergence criterion.
  - $\alpha$ indicates closeness of reference point from the Pareto-optimal front.
  - The value of running average of $\alpha$ over a prefixed number of generations to be close to *zero*.
  - Automatic and ensures an adequate convergence property.

# Test Setting

- We compare our concurrent-hybrid NSGA-II with serial-hybrid NSGA-II.

# Test Setting



- We compare our concurrent-hybrid NSGA-II with serial-hybrid NSGA-II.
- Test problems ranging from ZDT and DTLZ test suites and two practical problems: the welded beam design and the water resources planning problems.

# Test Setting



- We compare our concurrent-hybrid NSGA-II with serial-hybrid NSGA-II.
- Test problems ranging from ZDT and DTLZ test suites and two practical problems: the welded beam design and the water resources planning problems.
- Executed ten times with different seeds and best, median and worst values of performance metrics (function evaluations and hypervolume) noted.
- Termination criteria based on max function evaluations and error metric used.
- Diversity checked using hypervolume measure.

# Function Evaluation comparison



| Test Problem | Serial approach | | | Concurrent approach | | |
|---|---|---|---|---|---|---|
| | Best | Median | Worst | Best | Median | Worst |
| ZDT1 | 30,083 (0.9289) | 31,043 (0.9283) | 33,468 (0.9285) | **13,328** (0.9214) | 14,518 (0.9285) | 16,991 (0.9286) |
| ZDT2 | 29,384 (0.6526) | 31,760 (0.6530) | 32,344 (0.6532) | **1,861** (0.2100) | 13,748 (0.6513) | 15,716 (0.6510) |
| ZDT3 | 33,691 (0.7738) | 37,325 (0.7742) | 38,545 (0.7742) | **16,595** (0.7155) | 20,866 (0.7744) | 29,628 (0.7744) |
| ZDT4 | 35,006 (0.9274) | 54,214 (0.9284) | 63,584 (0.9286) | **34,459** (0.9286) | 37,724 (0.8982) | 43,142 (0.9286) |
| 3-DTLZ1 | 201,957 (1.664) | 252,952 (1.1965) | 351,954 (1.1964) | **66,369** (1.1995) | 146,506 (1.1931) | 290,792 (1.2002) |
| 3-DTLZ2 | 35,757 (0.8694) | 43,722 (0.8813) | 70,606 (0.8687) | **26,665** (0.8705) | 31,604 (0.8765) | 36,006 (0.8803) |
| 4-DTLZ2 | 69,449 (1.0861) | 93,835 (1.0701) | 128,794 (1.0750) | **61,028** (1.0960) | 74,187 (1.0834) | 194,581 (1.0782) |

# Function Evaluation Comparison- Exact Vs Approximate gradients



- Not obvious that in some real world engineering problems even such a high number is allowed.

| Test Problem | Exact gradient | | | Approximate gradient | | |
|---|---|---|---|---|---|---|
| | Best | Median | Worst | Best | Median | Worst |
| ZDT1 | **3,751** | 4,354 | 5,189 | 13,328 | 14,518 | 16,991 |
| ZDT2 | **1,706** | 4,510 | 5,721 | 1,861 | 13,748 | 15,716 |
| ZDT3 | **14,879** | 17,340 | 23,687 | 16,595 | 20,886 | 29,628 |
| ZDT4 | **18,763** | 21,975 | 26,148 | 34,459 | 37,724 | 43,142 |
| 3-DTLZ1 | **40,031** | 85,763 | 120,964 | 66,369 | 146,506 | 290,792 |
| 3-DTLZ2 | **15,017** | 19,230 | 24,380 | 26,665 | 31,604 | 36,006 |
| 4-DTLZ2 | **26,672** | 48,330 | 56,887 | 61,128 | 74,187 | 194,581 |

# Function Evaluation Comparison- Exact Vs Approximate gradients

- Not obvious that in some real world engineering problems even such a high number is allowed.

| Test Problem | Exact gradient | | | Approximate gradient | | |
|---|---|---|---|---|---|---|
| | Best | Median | Worst | Best | Median | Worst |
| ZDT1 | **3,751** | 4,354 | 5,189 | 13,328 | 14,518 | 16,991 |
| ZDT2 | **1,706** | 4,510 | 5,721 | 1,861 | 13,748 | 15,716 |
| ZDT3 | **14,879** | 17,340 | 23,687 | 16,595 | 20,886 | 29,628 |
| ZDT4 | **18,763** | 21,975 | 26,148 | 34,459 | 37,724 | 43,142 |
| 3-DTLZ1 | **40,031** | 85,763 | 120,964 | 66,369 | 146,506 | 290,792 |
| 3-DTLZ2 | **15,017** | 19,230 | 24,380 | 26,665 | 31,604 | 36,006 |
| 4-DTLZ2 | **26,672** | 48,330 | 56,887 | 61,128 | 74,187 | 194,581 |

- Drastic reduction in function evaluations.

# Diversity Comparison with Hypervolume

| Test Problem | Serial approach | | | Concurrent approach | | |
|---|---|---|---|---|---|---|
| | Best | Median | Worst | Best | Median | Worst |
| ZDT1 | **0.9291** | 0.9287 | 0.9283 | 0.9289 | 0.9276 | 0.9214 |
| ZDT2 | **0.6534** | 0.6530 | 0.6526 | 0.6531 | 0.6518 | 0.2100 |
| ZDT3 | 0.7743 | 0.7742 | 0.7738 | **0.7744** | 0.7737 | 0.7155 |
| ZDT4 | **0.9287** | 0.9286 | 0.9274 | **0.9287** | 0.9280 | 0.7758 |
| 3-DTLZ1 | 1.1981 | 1.1947 | 1.1664 | **1.2040** | 1.1994 | 1.1931 |
| 3-DTLZ2 | 0.8813 | 0.8694 | 0.8615 | **0.8850** | 0.8765 | 0.8645 |
| 4-DTLZ2 | 1.0983 | 1.0765 | 1.0602 | **1.0993** | 1.0857 | 1.0691 |
| WRP | 0.5703 | 0.5647 | 0.5635 | **0.5706** | 0.5660 | 0.5644 |
| WELD | 1.4196 | 1.4193 | 1.4082 | **1.4198** | 1.4188 | 1.4143 |

- HV values reached in 25,000 function evaluations for all test and practical problems.

# Slack variable ($\alpha$) as a Measure of Convergence

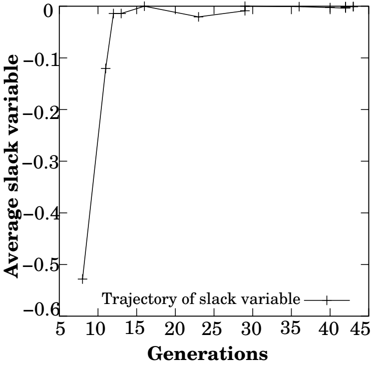


Figure: Variation of slack variable with generation in ZDT1.

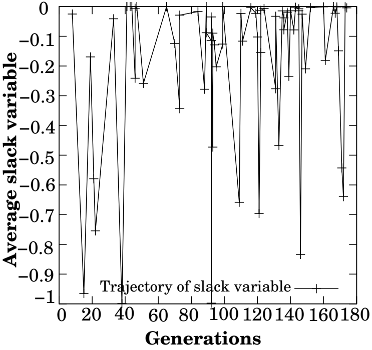


Figure: Variation of slack variable with generation in ZDT4.

# Effect of the Local Search on Convergence



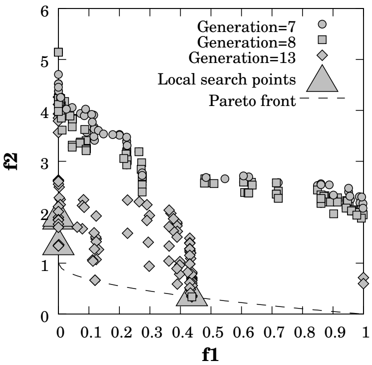


Figure: Populations approach the Pareto-optimal front faster in the concurrent-hybrid NSGA-II - ZDT1.

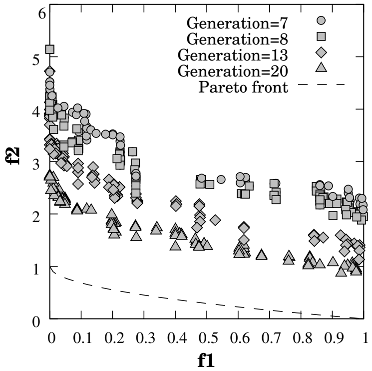


Figure: Populations approach the Pareto-optimal front slowly in the serial hybrid NSGA-II - ZDT1.

# Conclusion



- A concurrent-hybrid algorithm is proposed.

# Conclusion

- A concurrent-hybrid algorithm is proposed.
- Convergence objective achieved using ASF.

# Conclusion

- A concurrent-hybrid algorithm is proposed.
- Convergence objective achieved using ASF.
- Enhanced diversity preservation to be incorporated.

# Future Research Directions



- Steady state hybrid EMO.

# Future Research Directions

- Steady state hybrid EMO.
- Self adaptive $P_t^{local}$.

# Future Research Directions

- Steady state hybrid EMO.
- Self adaptive $P_t^{local}$.
- Clustering concurrent-hybrid NSGA-II.

# Thank You

- Next!!

# Thank You

- Next!!
- Tomi would provide you with ideas in generating an approximation of the points, which we have now generated.

# Thank You

- Next!!
- Tomi would provide you with ideas in generating an approximation of the points, which we have now generated.
- Questions ?